

Internal Coherence of Multi-Level AI-Generated Content: A Comparative Analysis of Novice and Experienced Teachers' Consistency within the IAT Framework

Hossein talebzadeh

talebzadehdr@gmail.com

Farhangian University <https://orcid.org/0009-0009-4312-1244>

Research Article

Keywords: Consistency Quotient (CQ), AI-TPACK, differentiated instruction, teacher professional development, generative AI, enacted pedagogical content knowledge

Posted Date: August 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10607265/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

This study introduces the Consistency Quotient (CQ) as a novel instrument to measure the internal coherence of AI-generated educational content. Through comparative content analysis of 356 samples from 89 teachers (38 novice, 51 experienced) across four differentiated models—Simplified, Enriched, Integrated Multi-Level, and Mixed-Ability—we examined how professional experience shapes coherence across conceptual, linguistic, and structural dimensions. Experienced teachers achieved significantly higher overall CQ scores ($M = 73.8$ vs. 67.4 , $d = 0.62$), with the largest advantages in conceptual ($d = 0.79$) and structural coherence ($d = 0.77$). Linguistic coherence showed a non-significant difference ($d = 0.38$), suggesting that structured prompt engineering effectively supports linguistic adaptation regardless of experience. Critically, negative case analysis revealed that 18.4% of novice teachers outperformed the experienced mean, while 11.8% of experienced teachers underperformed the novice mean, and overlap analysis indicated a 38.2% distribution overlap—challenging deterministic assumptions about professional experience. Pairwise comparisons identified the Simplified-Enriched pair as the most coherent and the Simplified-Mixed pair as the least coherent, highlighting the distinct challenges of vertical versus horizontal differentiation. These findings underscore that AI literacy and enacted pedagogical content knowledge are pivotal determinants of coherence, extending AI-TPACK frameworks by focusing on artifactual quality. The CQ offers a robust, practically applicable instrument for assessing teacher-AI collaboration and informing targeted professional development.

1. Introduction

The rapid proliferation of generative artificial intelligence (GenAI) has precipitated a paradigm shift in education, compelling a fundamental re-examination of pedagogical roles, assessment practices, and the nature of teacher professional knowledge (Selwyn, 2019; Luckin & Holmes, 2016; Cukurova, 2024). While GenAI tools offer unprecedented opportunities for personalizing learning, automating routine tasks, and augmenting teacher creativity (Dogan, 2025; Fernández-Sánchez et al., 2025; Chiu et al., 2023), their integration into classroom practice has been characterized by fragmentation and a persistent gap between technological availability and meaningful pedagogical application (Corbin et al., 2025; Ljungqvist et al., 2026). This disconnect has been described as a "jagged technological frontier" (Dell'Acqua et al., 2026, p. 405), where AI assistance improves performance for some tasks while undermining it for others. The consequences are particularly acute in educational settings, where teachers' professional judgment, contextual awareness, and pedagogical reasoning remain irreplaceable (Cukurova, 2025; Mishra & Bozkurt, 2026; Selwyn, 2021).

Central to the challenge of AI integration is the transformation of the teacher's role from content deliverer to learning designer and pedagogical orchestrator (Laurillard, 2024; Bosch et al., 2025). This shift requires teachers to develop not only technological proficiency but also what scholars have termed "AI literacy" (Chiu et al., 2024; Heung et al., 2026) and "AI-TPACK" (An et al., 2026; Chen et al., 2026; Sun et al., 2026)—the ability to integrate AI tools with content knowledge and pedagogical strategies in conceptually coherent ways. However, existing research has predominantly focused on the *efficiency*

gains afforded by AI rather than on the *quality* and *internal coherence* of the educational artifacts teachers produce with AI assistance (Cao et al., 2026; Chan & Lee, 2025; Samuel et al., 2025). This emphasis on efficiency over quality risks perpetuating what Selwyn (2019, 2021) has critiqued as a techno-solutionist discourse, wherein technological capability is conflated with educational value, and the complex, situated dimensions of teaching are reduced to algorithmic optimization (Ljungqvist et al., 2026; Cukurova, 2025).

A critical yet underexplored dimension of AI-augmented teaching is the **internal coherence** of multi-level content that teachers generate for diverse learners. In differentiated instruction—a pedagogical approach that has gained renewed urgency in the context of AI-mediated personalization (Vanhees et al., 2025; Afidah & Hernawan, 2025)—teachers are expected to produce multiple versions of the same lesson content tailored to students with varying readiness levels (Tomlinson, 2014). The four content models examined in this study—Simplified, Enriched, Integrated Multi-Level, and Mixed-Ability Group activities—represent the full spectrum of differentiation practices, ranging from individual level-appropriate materials to collaborative tiered projects (Talebzadeh, 2025; Rosyidi et al., 2025). Across these representations, teachers must preserve the conceptual core, linguistic consistency, and structural integrity of the subject matter. This capacity to maintain coherence is not merely a technical skill but a manifestation of deeper pedagogical content knowledge (Shulman, 1986, 1987), pedagogical reasoning (Loughran et al., 2016), and what Carlson et al. (2019) have termed "enacted PCK" within the Refined Consensus Model. The extent to which teachers can sustain this coherence when collaborating with GenAI tools is a pressing empirical question, particularly as novice and experienced teachers may differ significantly in their ability to critically evaluate and integrate AI-generated outputs (Dinçer, 2026; Wang, 2026; Hobbs, 2020).

Despite the growing body of research on AI in education, several critical gaps persist. First, most studies have examined AI-generated content in isolation, without systematically comparing the coherence of content produced by teachers with varying professional experience (Chiu et al., 2023; Zawacki-Richter et al., 2025). Second, existing frameworks for assessing teacher AI competence (Chiu et al., 2024; An et al., 2026; Zhou et al., 2025) have focused on knowledge domains rather than on the quality of educational artifacts produced through human-AI collaboration. Third, while differentiation is widely advocated, there is a lack of robust instruments to measure the internal coherence of multi-level content across conceptual, linguistic, and structural dimensions (Braun & Clarke, 2023; Messner et al., 2025). Fourth, the role of professional experience in shaping teachers' capacity to maintain coherence when working with AI has received limited attention, with contradictory findings emerging from studies comparing novice and experienced teachers' interactions with AI tools (Dinçer, 2026; Cao et al., 2026; Wang, 2026).

To address these gaps, this study introduces the **Consistency Quotient (CQ)**, a composite index designed to measure the internal coherence of four content models produced by a single teacher for the same lesson topic. Crucially, the CQ assesses the teacher's **enacted pedagogical content knowledge (enacted PCK)** as manifested in the final, teacher-edited and teacher-approved outputs, rather than the raw outputs of AI tools alone. All content generation followed a teacher-in-the-loop workflow: teachers

crafted structured prompts, received initial AI-generated drafts, and engaged in substantive editing, revision, and validation to ensure alignment with their pedagogical intentions. The final submitted models represent the teacher's curated artifacts, positioning CQ as a measure of the teacher's capacity to critically evaluate, adapt, and integrate AI-generated content into coherent, differentiated instructional materials. This operationalization aligns with contemporary conceptualizations of enacted PCK (Carlson et al., 2019) and teacher agency (Priestley et al., 2015, 2026), wherein professional knowledge is performed and materialized in practice.

The CQ operationalizes coherence across three dimensions: Conceptual Coherence—the preservation of the conceptual core across all models, informed by Shulman's (1986, 1987) PCK framework and Bernstein's (1996, 2000) theory of pedagogic discourse; Linguistic Coherence—the consistency of vocabulary, terminology, and language register, drawing on Mayer's (2024) cognitive theory of multimedia learning; and Structural Coherence—the consistency in overall content structure, operationalized through van Es and Sherin's (2026) conceptualization of teacher noticing and shaping practices.

Employing a comparative content analysis with a hybrid quantitative-qualitative approach, this study analyzes 356 content samples generated by 89 teachers across two broad experience categories: Novice (N = 38) and Experienced (N = 51). Each teacher produced four content models for a single lesson topic. The study addresses three research questions: (1) What patterns of CQ emerge across the four content models generated by novice vs. experienced teachers? (2) Is there a significant difference in CQ between novice and experienced teachers, and in which coherence dimension is this difference most pronounced? (3) In which pair of models is the greatest and least coherence observed, and how does this pattern differ between the two groups?

This study contributes to the literature on AI-augmented teaching in three significant ways. First, it introduces the CQ as a novel instrument for assessing the quality of teacher-AI collaboration, moving beyond efficiency metrics to capture pedagogical integrity. Second, it provides empirical evidence on how professional experience shapes teachers' capacity to maintain coherence when generating differentiated content with GenAI. Third, it extends existing AI-TPACK frameworks by focusing on the artifactual quality of teacher-edited AI outputs. To ensure validity and trustworthiness, the study incorporates a four-part reflexive structure—Alternative Hypotheses Examination, Negative Case Analysis, Overlap Analysis, and Self-Critique—designed to mitigate confirmation bias and enhance scholarly rigor.

2. Theoretical Framework

To establish the theoretical foundations for the Consistency Quotient (CQ) and to contextualize the empirical investigation of teacher-AI collaboration, this section synthesizes three interconnected bodies of literature. First, it traces the evolution of teacher professional knowledge from Pedagogical Content Knowledge (PCK) to the contemporary AI-TPACK frameworks, establishing the theoretical grounding for

conceptualizing teacher expertise in AI-augmented contexts. Second, it articulates a multi-dimensional theory of coherence, drawing on distinct theoretical traditions to operationalize the conceptual, linguistic, and structural dimensions of content consistency, and provides a theoretical justification for the equal weighting of these dimensions in the CQ composite index. Third, it situates the study within the Integrated AI Triad (IAT) model as a guiding analytical lens for understanding the systemic redefinition of teacher roles. Finally, in accordance with the study's reflexive commitment to mitigating confirmation bias, this section explicitly examines alternative hypotheses that might explain why novice teachers could theoretically demonstrate higher coherence than their experienced counterparts.

2.1. The Evolution of Teacher Professional Knowledge: From PCK to AI-TPACK

The conceptualization of teacher professional knowledge has undergone significant theoretical development over the past four decades, reflecting the changing demands of educational practice and the increasing complexity of instructional contexts. Shulman's (1986, 1987) foundational work on Pedagogical Content Knowledge (PCK) represented a paradigm shift by arguing that effective teaching requires more than content expertise or general pedagogical skills; rather, it demands the integration of content and pedagogy into a specialized form of knowledge that enables teachers to transform subject matter into forms that are intellectually accessible to learners. Shulman (1987) identified seven categories of teacher knowledge, with PCK occupying a central position as the distinctive contribution of the teaching profession—the capacity to represent and formulate subject matter in ways that make it comprehensible to others, to anticipate student misconceptions, and to adapt instruction to diverse learner needs.

The advent of digital technologies prompted an extension of this framework, culminating in the Technological Pedagogical Content Knowledge (TPACK) model, which conceptualizes effective technology integration as the intersection of technological, pedagogical, and content knowledge (Koehler & Mishra, 2009). The TPACK framework has been extensively applied in research on teacher technology integration, emphasizing that meaningful use of technology requires teachers to understand not only how to use specific tools but also how technology interacts with pedagogical strategies and content representations. However, the rapid emergence of generative AI tools has revealed limitations in the TPACK model's capacity to address the unique challenges posed by AI. As scholars have noted, AI tools differ fundamentally from previous educational technologies in their generative capacity, their autonomy, and their potential to perform tasks traditionally reserved for teachers, such as content generation, assessment, and even pedagogical decision-making (Cukurova, 2025; Mishra & Bozkurt, 2026).

In response, recent scholarship has proposed extensions to the TPACK framework specifically addressing AI integration. The AI-TPACK model (An et al., 2026; Sun et al., 2026) distinguishes between the knowledge required for human teachers and that applicable to AI machine teachers, recognizing that AI can function not merely as a tool but as a semi-autonomous agent in the instructional process.

Similarly, the I-TPACK framework (Chen et al., 2026) emphasizes the integration of AI knowledge with content and pedagogy, identifying discipline-sensitive patterns in how teachers enact AI-TPACK domains. Critically, scholars have highlighted that AI-Content Knowledge is enacted through practices of validation, contextualization, and critical evaluation rather than content generation alone, underscoring the central role of teachers' epistemic and pedagogical judgment (Chen et al., 2026; Mishra & Bozkurt, 2026).

Despite these advances, existing AI-TPACK frameworks have predominantly focused on the *knowledge* teachers possess—the declarative and procedural knowledge required to use AI tools—rather than on the *quality of educational artifacts* teachers produce through human-AI collaboration. This gap is significant because teacher knowledge is ultimately manifested in practice, in what Carlson et al. (2019) have termed "enacted PCK" within the Refined Consensus Model (RCM) of PCK. The RCM distinguishes among three realms of PCK: collective PCK (the knowledge held by the professional community), personal PCK (the knowledge an individual teacher brings to a situation), and enacted PCK (the knowledge that is mobilized and manifested in the act of teaching). It is this enacted PCK that is most directly relevant to understanding how teachers translate their knowledge into coherent instructional materials. The present study addresses this gap by introducing the Consistency Quotient as a measure of enacted PCK in the context of AI-augmented content generation, thereby extending AI-TPACK frameworks to encompass the artifactual quality of teacher-AI collaboration.

2.2. Conceptualizing Coherence in Differentiated Educational Content: A Multi-Dimensional Framework

The concept of coherence has been central to educational theory, particularly in discussions of curriculum design, instructional quality, and teacher knowledge. Coherence refers to the logical, conceptual, and linguistic consistency that enables learners to construct meaningful understanding from instructional materials (Vanhees et al., 2025; Mayer, 2024). In the context of differentiated instruction—the practice of tailoring content, process, and product to students' varying readiness levels, interests, and learning profiles (Tomlinson, 2014)—coherence poses a particular challenge: teachers must produce multiple representations of the same content while ensuring that the core conceptual structure remains intact across all versions. This study operationalizes coherence across three theoretically grounded dimensions: Conceptual, Linguistic, and Structural.

The relationships among these dimensions and their integration into the overall CQ score are illustrated in Fig. 1.

Conceptual Coherence refers to the preservation of the conceptual core—the key definitions, principles, and causal relationships that constitute the essential structure of a disciplinary domain—across all content models. This dimension draws on Bernstein's (1996, 2000) theory of pedagogic discourse and the distinction between vertical and horizontal discourse. Vertical discourse refers to the systematic, hierarchically organized knowledge structures characteristic of academic disciplines, while horizontal discourse refers to localized, context-dependent knowledge. Bernstein argued that the

recontextualization of vertical discourse into pedagogic forms requires careful attention to maintaining the integrity of the conceptual structure while making it accessible to learners. In the context of AI-generated content, conceptual coherence requires that the essential disciplinary structure be preserved across simplified, enriched, and integrated versions, such that the "conceptual core" remains identifiable and intact (Talebzadeh, 2025; Rosyidi et al., 2025). This aligns with Shulman's (1986, 1987) emphasis on the importance of teachers' ability to transform content knowledge while preserving its intellectual integrity.

Linguistic Coherence concerns the consistency of vocabulary, terminology, and language register across content models, appropriately adjusted for each target audience. This dimension draws on Mayer's (2024) Cognitive Theory of Multimedia Learning, which emphasizes that learners process information through dual channels (verbal and visual) and that meaningful learning requires the coherent integration of verbal and visual representations. Mayer's (2024) fifteen evidence-based principles of multimedia design highlight the importance of reducing extraneous cognitive load and ensuring that language is clear, concise, and tailored to learners' prior knowledge. In the context of differentiated content, linguistic coherence requires that although vocabulary complexity and register may be adjusted for different audiences (e.g., simplified language for struggling learners, more sophisticated terminology for advanced students), the core terminology and conceptual labels remain consistent across versions to prevent confusion and support the development of disciplinary literacy (Vanhees et al., 2025).

Structural Coherence refers to the consistency in overall content structure—the logical flow, sequencing of ideas, causal relationships, and argumentation chains that organize the content. This dimension is operationalized through van Es and Sherin's (2026) conceptualization of teacher noticing and shaping practices. Their research on teacher noticing identifies three categories of shaping practices that teachers employ to advance their in-the-moment instructional decisions: eliciting details (drawing out specific aspects of student thinking), leveraging participant structures (using collaborative arrangements to support learning), and creating and using material and social resources (designing and adapting instructional materials to support understanding). In the context of content generation, structural coherence requires that the logical organization of the content—the sequence in which concepts are introduced, the relationships between ideas, and the progression from simple to complex—remains consistent across differentiated versions, even as the depth, complexity, and language adjust to meet diverse learner needs. This dimension is also informed by pedagogical reasoning (Loughran et al., 2016), which emphasizes the importance of teachers' ability to justify and articulate the structural choices they make in designing instruction.

Theoretical Justification for Equal Weighting in the Consistency Quotient. A critical methodological decision in the construction of the CQ composite index is the assignment of equal weight (33.3%) to each of the three coherence dimensions. This equal weighting is not arbitrary but is grounded in theoretical considerations regarding the nature of differentiated content and the role of coherence in supporting meaningful learning. In the context of multi-level content generation, each dimension represents a distinct yet equally essential facet of pedagogical quality. Conceptual coherence ensures

that the disciplinary core is preserved across all versions, thereby enabling students at different readiness levels to develop a shared understanding of foundational concepts (Bernstein, 1996, 2000; Shulman, 1986). Linguistic coherence ensures that students can navigate across versions without encountering inconsistent terminology or confusing referents, thereby supporting the development of disciplinary literacy (Mayer, 2024; Vanhees et al., 2025). Structural coherence ensures that the logical flow and argumentation chains are preserved, enabling students to follow the progression of ideas and develop coherent mental models (van Es & Sherin, 2026; Loughran et al., 2016). A deficiency in any single dimension—whether the fragmentation of key definitions (conceptual), the introduction of inconsistent terminology (linguistic), or the disruption of causal argumentation (structural)—can independently undermine the overall coherence of the instructional materials and lead to learner confusion. This insight is consistent with cognitive load theory (Sweller, 1988; Mayer, 2024), which posits that incoherence in any aspect of instructional design imposes extraneous cognitive load that impedes learning. Therefore, assigning equal weight to the three dimensions as a baseline weighting (baseline weighting) is theoretically justified: it reflects the premise that each dimension contributes independently and equally to the overall coherence of differentiated content, and that deficits in any dimension have the potential to compromise the integrity of the instructional design. While future research may investigate differential weighting based on subject domain or learner characteristics, the equal weighting serves as a theoretically defensible baseline for the present exploratory study.

2.3. The Integrated AI Triad (IAT) as a Guiding Analytical Lens

This study utilizes the Integrated AI Triad (IAT) model (Talebzadeh, 2025) as a guiding analytical lens for understanding the reconfiguration of educational roles in the age of AI. The IAT model positions AI as a mediator among three core pillars: Teachers, Managers, and Families, each of whom must develop what the framework terms "bilingual competence"—fluency in both human pedagogical traditions and AI technologies. In this study, the IAT framework serves as a conceptual tool for understanding the transformation of the teacher's role from a content deliverer to an AI-Integrated Learning Design Partner (AI-LDP). This role encompasses four key competencies: smart task design (creating scenario-based and creative tasks resistant to AI plagiarism), differentiated content generation (producing multi-level content for diverse learner needs, consistent with the differentiation framework employed in this study), assessment automation (using AI to grade objective assessments), and rubric design (constructing analytical rubrics for higher-order thinking).

Crucially, the IAT framework informed the design of the professional development workshop that generated the data for this study. The workshop protocol, grounded in the IAT's pedagogical principles, provided teachers with structured prompt engineering templates (Role + Audience + Format) for generating differentiated content across the four model types examined in this research. Moreover, the workshop emphasized a teacher-in-the-loop (human-in-the-loop) workflow, requiring teachers to edit, revise, and validate AI-generated drafts to ensure alignment with their pedagogical intentions, curricular standards, and students' needs. This design ensures that the content models analyzed in this study

reflect the teachers' enacted professional knowledge (enacted PCK) rather than unmediated AI output, thereby positioning CQ as a valid measure of teacher competence in AI-augmented instructional design.

2.4. Alternative Hypotheses Examination: Why Novice Teachers Might Demonstrate Higher Coherence

In accordance with the study's commitment to reflexive rigor and the mitigation of confirmation bias, this subsection examines alternative theoretical explanations for why novice teachers might, counterintuitively, demonstrate higher Consistency Quotients than their experienced counterparts. While conventional wisdom might predict that experienced teachers, with their deeper content knowledge and pedagogical expertise, would produce more coherent multi-level content, several theoretically grounded counterarguments warrant careful consideration.

First, novice teachers may possess more recent and extensive formal training in contemporary pedagogical approaches, including differentiated instruction, constructivist learning theories, and student-centered pedagogies (Day et al., 2006; Pillen et al., 2013). Their preservice education has likely emphasized the importance of maintaining conceptual coherence across differentiated materials and may have included explicit instruction in designing for diverse learners. In contrast, experienced teachers' training may predate the widespread emphasis on differentiation and may have been grounded in more transmission-oriented pedagogical models, potentially leading to less systematic attention to coherence across differentiated content (Vanhees et al., 2025; Afidah & Hernawan, 2025).

Second, novice teachers may adhere more closely to curriculum ideals and instructional templates without the idiosyncratic modifications that experienced teachers develop over years of practice (Priestley et al., 2015, 2026). While experienced teachers' deep knowledge of their students, school context, and curriculum constraints enables them to adapt content in ways that novice teachers cannot, these adaptations may sometimes introduce inconsistencies that disrupt the conceptual coherence of the material. Experienced teachers may be more likely to "over-edit" AI-generated outputs, introducing their own preferred examples, explanations, and organizational structures that, while contextually responsive, may fragment the underlying conceptual structure (Ljungqvist et al., 2026; Wang, 2026).

Third, experienced teachers may be more resistant to AI-generated suggestions that conflict with their established pedagogical schemas, leading to cognitive dissonance (Festinger, 1957; Kaswa, 2025) and editorial decisions that prioritize alignment with personal beliefs over maintaining coherence with other content versions. Mishra and Bozkurt's (2026) characterization of AI as a "smart drunk intern"—highly capable but fundamentally unreliable—captures this tension: experienced teachers, aware of the limitations of AI, may be more critical and selective in their acceptance and adaptation of AI-generated content, potentially introducing greater variability across content versions.

Fourth, novice teachers' professional identity is more fluid and potentially more adaptable to new technologies and pedagogical approaches (Pillen et al., 2013; Gao et al., 2025). Day et al. (2006) found that teacher identities may be more or less stable at different times, and that early career teachers often

demonstrate greater openness to innovation and change. This identity fluidity may enable novice teachers to engage more flexibly with AI tools, accepting and integrating AI-generated suggestions with less resistance, thereby producing more consistent outputs across content versions.

Fifth, the ecological conditions that shape teacher agency—including the interplay between individual capacity, contextual constraints, and structural opportunities—may differentially affect novice and experienced teachers (Priestley et al., 2015, 2026; Biesta et al., 2015). Novice teachers, unencumbered by the accumulated professional routines and institutional expectations that may constrain experienced teachers' practice, may exercise greater agency in adopting new approaches to content generation, resulting in more systematic attention to coherence across differentiated materials.

These alternative hypotheses serve as critical theoretical counterweights to the assumption that experienced teachers will uniformly outperform novices in maintaining coherence. They underscore the importance of empirical investigation and ensure that the study's analysis remains open to findings that contradict conventional expectations, thereby strengthening the validity and scholarly contribution of the research.

3. Methodology

This section delineates the methodological framework employed to investigate the internal coherence of AI-generated educational content produced by novice and experienced teachers. The section details the research design, participant characteristics and context, the content generation workflow, the coding framework for assessing coherence dimensions, the calculation of the Consistency Quotient (CQ), procedures for ensuring inter-coder reliability, and the analytical strategies for addressing the research questions. Throughout, the methodology is designed to support the study's reflexive commitment to rigor, transparency, and mitigation of confirmation bias.

3.1. Research Design and Context

This study employs a comparative content analysis with a hybrid quantitative-qualitative approach, situated within a professional development workshop series based on the Integrated AI Triad (IAT) model (Talebzadeh, 2025). The research design is exploratory and inductive, consistent with the study's aim to identify emergent patterns in the coherence of AI-generated content across teacher experience levels (Braun & Clarke, 2023). The workshop, titled "Bilingual Teacher Series: Personalized Content Creation," equipped teachers with competencies in using GenAI tools following structured prompt engineering templates (Role + Audience + Format). Crucially, the workshop emphasized a teacher-in-the-loop workflow: teachers crafted prompts, received AI-generated drafts, and engaged in substantive editing and validation before submitting final artifacts. This ensured that content models reflected teachers' enacted pedagogical content knowledge (Carlson et al., 2019) rather than unmediated AI output.

3.2. Participants

Participants comprised 89 teachers from Iranian educational contexts, recruited through professional networks and institutional partnerships. The sample was stratified into two broad experience categories:

- **Novice Teachers (N = 38):** Recent graduates who had completed teacher education programs but had not yet begun formal teaching practice (two cohorts: Novice A, N = 19; Novice B, N = 19).
- **Experienced Teachers (N = 51):** In-service teachers with two or more years of full-time teaching experience (two cohorts: Experienced A, N = 25; Experienced B, N = 26).

Participants represented diverse subject areas (mathematics, science, language arts, social studies, humanities). All provided informed consent, with ethical protocols ensuring confidentiality and acknowledgment of the researcher's dual role as IAT developer and workshop instructor.

3.3. Content Generation Procedure

Each teacher generated four distinct content models for a single self-selected lesson topic, following structured templates representing key differentiation strategies:

1. **Simplified Content (Template 1):** Adapted for struggling learners with simplified vocabulary, reduced complexity, concrete examples, and scaffolded explanations.
2. **Enriched Content (Template 2):** Adapted for advanced/gifted students with sophisticated terminology, extended depth, complex examples, and challenging extension questions.
3. **Integrated Multi-Level Activity (Template 4):** A tiered group project in which three groups (basic, intermediate, advanced) work on differentiated tasks culminating in a shared product.
4. **Mixed-Ability Group Activity (Template 5):** A collaborative project for mixed-ability groups with distinct roles and differentiated responsibilities.

Teachers followed a structured workflow: (a) crafted detailed prompts using the IAT-based template, (b) received initial AI-generated drafts, (c) engaged in substantive editing and validation, and (d) submitted final teacher-edited versions. This process ensured content models represented teachers' enacted PCK and professional judgment. The complete dataset comprised 356 content samples (89 teachers × 4 models).

3.4. Coding Framework and the Consistency Quotient (CQ)

The Consistency Quotient (CQ) was operationalized as a composite index ranging from 0 to 100, aggregating scores across three theoretically grounded coherence dimensions, each weighted equally at 33.3%. The coding framework was developed iteratively through deductive (theory-driven) and inductive (data-driven) processes, drawing on theoretical foundations articulated in Section 2.2 and refined through pilot coding.

Conceptual Coherence (33.3%) assessed the preservation of the "conceptual core"—key definitions, principles, and causal relationships—across the four models. Coders evaluated whether core concepts appeared consistently, definitions remained stable (despite simplification or enrichment), and essential

relationships were preserved. This dimension drew on Bernstein's (1996, 2000) theory of pedagogic discourse and Shulman's (1986, 1987) emphasis on content knowledge integrity.

Linguistic Coherence (33.3%) assessed the consistency of vocabulary, terminology, and language register across models, appropriately adjusted for each audience. Coders evaluated whether key terminology remained consistent across versions, whether register shifts were appropriate and justified, and whether linguistic adjustments supported development of disciplinary literacy (Mayer, 2024; Vanhees et al., 2025).

Structural Coherence (33.3%) assessed the consistency of overall content structure—logical flow, sequencing of ideas, causal relationships, and argumentation chains. Coders evaluated whether concept introduction sequences remained consistent, argumentation chains were preserved, and organizational logic was identifiable across versions. This dimension was operationalized through van Es and Sherin's (2026) conceptualization of teacher noticing and shaping practices.

For each dimension, coders assigned scores on a 0–100 scale based on coherence observed across the four models. Detailed rubrics specified criteria for each score band (e.g., 0–25: minimal coherence; 26–50: partial coherence with significant inconsistencies; 51–75: moderate coherence with minor inconsistencies; 76–100: high coherence with consistent preservation across all models). The final CQ for each teacher was calculated as:

$$CQ = \frac{C_{\text{conceptual}} + C_{\text{linguistic}} + C_{\text{structural}}}{3}$$

3.5. Inter-Coder Reliability and Quality Assurance

To ensure rigor and trustworthiness, two coders with advanced degrees in education participated in the analysis. Both received intensive training in the coding framework, including practice sessions with sample data not included in the final dataset. A double-coding protocol was implemented for 20% of the teacher output sets (n = 18 teachers, representing 72 individual content samples), randomly selected across teacher cohorts and model types. Cohen's kappa coefficient was calculated for each coherence dimension and for the overall CQ score, with a minimum threshold of $\kappa \geq 0.80$ established as acceptable reliability (Braun & Clarke, 2023; Messner et al., 2025). Disagreements were resolved through consensus discussions, and final coding decisions for the double-coded subset were incorporated once consensus was reached.

Additional quality assurance measures included: (a) regular calibration meetings between coders throughout the process, (b) systematic documentation of coding decisions and rationales, (c) maintenance of a coding audit trail, and (d) periodic cross-checking of coding consistency across cohorts and model types. Coding was conducted using ATLAS—ti qualitative data analysis software.

3.6. Reflexive Measures to Mitigate Confirmation Bias

Four reflexive measures were systematically integrated into the research design to mitigate confirmation bias and enhance scholarly rigor (Braun & Clarke, 2023; Messner et al., 2025; Ritzhaupt et al., 2026):

First, Alternative Hypotheses Examination (detailed in Section 2.4): The theoretical framework explicitly examined why novice teachers might theoretically demonstrate higher coherence than experienced teachers, ensuring the analysis was not biased toward confirming a conventional "experience advantage" hypothesis.

Second, Negative Case Analysis

The findings section systematically identifies (a) novice teachers with CQ scores above the experienced teacher mean and (b) experienced teachers with CQ scores below the novice teacher mean. For each exceptional case, qualitative analysis characterizes factors explaining the deviation.

Third, Overlap Analysis

The study calculates and reports the percentage of overlap between novice and experienced teacher CQ distributions. Following methodological guidelines (Howard, 1980), if overlap exceeds 30–40%, practical significance of group differences is considered limited, regardless of statistical significance.

Fourth, Self-Critique and Reflexivity (detailed in Section 5.2): The discussion includes a subsection explicitly acknowledging how selecting "coherence" as the primary dimension may favor experienced teachers, how the researcher's IAT commitment may have influenced analysis, how AI-assisted analysis may have introduced biases, and specific mitigation steps taken.

3.7. Analytical Strategy

The analysis proceeded in three phases, corresponding to the three research questions.

Phase 1: Descriptive and Comparative Analysis. For each coherence dimension and overall CQ, descriptive statistics (mean, median, SD, min, max, percentiles) were calculated separately for novice and experienced teachers. Histograms and boxplots visualized score distributions. Effect sizes (Cohen's *d*) quantified group differences, and independent-samples *t*-tests (or non-parametric equivalents) assessed statistical significance.

Phase 2: Negative Case and Overlap Analysis. Following protocols in Section 3.6, negative cases were systematically identified and subjected to qualitative analysis. Overlap percentages were calculated using Howard's (1980) approach, comparing proportions of scores in overlapping regions of the two distributions for overall CQ and each dimension separately.

Phase 3: Pairwise Model Coherence Comparison. Coherence scores were compared across the four model types for each teacher. Pairwise comparisons (e.g., Simplified-Enriched, Integrated-Mixed) identified model pairs with greatest and least coherence. This analysis was conducted separately for

novice and experienced groups, enabling identification of differential patterns. Statistical testing employed repeated-measures ANOVA (or non-parametric equivalents).

All statistical analyses were conducted using SPSS (Version 28) and R (Version 4.2), with statistical significance set at $p < 0.05$. Qualitative data were analyzed using thematic analysis (Braun & Clarke, 2020) supported by ATLAS.ti, with themes identified through iterative reading, coding, and refinement.

3.8.Ethical Considerations

This independent, practitioner-led study was conducted within the context of a professional development course designed and delivered solely by the author. The research involved standard educational practices and presented no more than minimal risk to participants. Consequently, formal approval from a named institutional ethics committee was not required. Nonetheless, all procedures strictly adhered to the ethical principles outlined in the Declaration of Helsinki, with careful attention to participant confidentiality, anonymity, and voluntary participation.

Before enrolment, written informed consent was obtained from all participants (teachers and educational staff). Each participant was fully informed of the study's objectives, data collection methods, the scope of the intervention, and their unconditional right to withdraw at any time without negative consequences. All data were anonymized before analysis to ensure privacy and confidentiality.

4. Findings

This section presents the findings of the comparative content analysis of 356 content samples generated by 89 teachers (38 novice, 51 experienced) across four content models per teacher. The findings are organized according to the three research questions, beginning with descriptive statistics and group comparisons, followed by negative case analysis, overlap analysis, and pairwise model coherence comparisons. Throughout, the findings are presented with attention to both quantitative patterns and qualitative nuances that illuminate the nature of coherence—and its absence—in AI-generated educational content.

4.1. Descriptive Statistics and Group Differences

4.1.1. Overall Consistency Quotient (CQ)

Table 2 presents the descriptive statistics for the overall Consistency Quotient (CQ) across novice and experienced teacher groups, along with effect sizes.

Table 2
Descriptive Statistics of Consistency Quotient (CQ) with Effect Sizes

Measure	Novice Teachers (N = 38)	Experienced Teachers (N = 51)	Effect Size (Cohen's d)
Mean CQ	67.4	73.8	0.62
Median CQ	68.1	74.5	—
SD	8.9	9.2	—
Minimum	48.2	51.7	—
Maximum	84.6	91.3	—
25th Percentile	61.3	67.2	—
75th Percentile	73.8	80.1	—

Independent-samples t-test revealed a statistically significant difference between the two groups, with experienced teachers demonstrating higher overall CQ scores ($M = 73.8$, $SD = 9.2$) compared to novice teachers ($M = 67.4$, $SD = 8.9$), $t(87) = 3.41$, $p < .001$. The effect size (Cohen's $d = 0.62$) indicates a moderate to large practical significance.

Figure 2. Distribution of Consistency Quotient (CQ) scores for novice ($N = 38$) and experienced ($N = 51$) teachers. Vertical dashed lines indicate group means ($M_{\text{novice}} = 67.4$, $M_{\text{exp}} = 73.8$). The overlap between distributions is 38.2%. The difference is statistically significant ($t(87) = 3.41$, $p < .001$, Cohen's $d = 0.62$).

To address the second research question—whether the difference between groups is most pronounced in a specific coherence dimension—the three sub-dimensions were analyzed separately.

Table 2
a: Dimensional Coherence Scores by Teacher Group

Coherence Dimension	Novice Teachers (N = 38)	Experienced Teachers (N = 51)	t-value	p-value	Cohen's d
Conceptual Coherence	69.1 (9.4)	76.2 (8.7)	3.72	< .001	0.79
Linguistic Coherence	68.7 (10.1)	72.5 (9.8)	1.81	.074	0.38
Structural Coherence	64.4 (11.2)	72.7 (10.3)	3.58	< .001	0.77

Note

Values are presented as Mean (SD).

The dimensional analysis reveals a nuanced pattern. The largest difference between novice and experienced teachers was observed in **Conceptual Coherence** ($d = 0.79$), followed closely by **Structural Coherence** ($d = 0.77$). Both differences were statistically significant at $p < .001$. In contrast, **Linguistic Coherence** showed a smaller and non-statistically significant difference ($d = 0.38, p = .074$).

Qualitative Observations:

Conceptual Coherence — Experienced teachers demonstrated superior preservation of the conceptual core, particularly in abstract topics such as "Earth's Layers" and "Constitutionalism," where they maintained consistent definitions and principles across all four models while skillfully adjusting depth. Novice teachers showed stronger conceptual coherence in mathematics and science topics with explicit, formulaic structures (e.g., "Fractions") but occasionally introduced inconsistencies in more abstract subjects.

Structural Coherence — Experienced teachers consistently maintained the same logical progression (e.g., "observation → hypothesis → evidence → conclusion" in science; "causes → events → consequences" in history) across all models, adjusting explanation depth at each step. Novice teachers exhibited greater variability, sometimes omitting structural components in simplified versions—for example, eliminating the causal chain between isotope abundance and average atomic mass in the "Atomic Mass" lesson.

Linguistic Coherence — Both groups demonstrated comparable linguistic adaptation, likely due to the structured prompt engineering templates provided in the workshop. However, experienced teachers showed slightly more sophisticated terminology consistency, maintaining key terms (e.g., "physical change," "chemical change") across all models, while novices sometimes used inconsistent terminology across versions.

4.2. Negative Case Analysis: Exceptional Cases Contradicting the Main Pattern

In accordance with the study's reflexive commitment to identifying and analyzing cases that contradict the main findings, Table 3 summarizes the negative cases identified.

Table 3
Negative Case Analysis — Summary of Exceptional Cases

Category	Cases Identified	Mean CQ	Range
Novice teachers above experienced mean (CQ > 73.8)	7 of 38 (18.4%)	79.2	74.1–84.6
Experienced teachers below novice mean (CQ < 67.4)	6 of 51 (11.8%)	61.3	51.7–66.9

Exceptional Novice Teachers: The seven novice teachers who outperformed the experienced mean shared three characteristics: strong subject matter knowledge in their selected topics (e.g., a physics graduate achieving CQ = 84.6 on "Free Fall"), systematic adherence to the "Role + Audience + Format" prompt engineering framework with detailed contextual elaboration, and active editing of AI outputs rather than uncritical acceptance.

Low-Performing Experienced Teachers

The six experienced teachers who scored below the novice mean exhibited "over-editing" of AI-generated content, introducing idiosyncratic examples and analogies that fragmented coherence. They also showed resistance to AI suggestions, rejecting conceptually coherent AI outputs in favor of their own pedagogical preferences. Several reported significant time constraints, completing their models more quickly and with less revision.

4.3. Overlap Analysis

Table 4
Overlap Analysis – Distribution Overlap by Coherence Dimension

Coherence Dimension	Overlap Percentage	Interpretation
Overall CQ	38.2%	Moderate overlap; groups distinguishable but share substantial common ground
Conceptual Coherence	32.7%	Lower overlap; meaningful distinction between groups
Linguistic Coherence	48.6%	Higher overlap; limited practical distinction
Structural Coherence	33.1%	Lower overlap; meaningful distinction between groups

The overall CQ overlap of 38.2% suggests that while novice and experienced teachers are distinguishable groups, there is substantial common ground in their ability to maintain coherence. Conceptual Coherence (32.7%) and Structural Coherence (33.1%) showed the lowest overlap, confirming these dimensions as where professional experience has the greatest practical impact. Linguistic Coherence (48.6%) showed the highest overlap, suggesting linguistic adaptation is less dependent on experience and more accessible to all teachers, likely due to structured prompt support.

4.4. Pairwise Model Coherence Comparison

Table 5
Pairwise Model Coherence Comparison

Model Pair	Novice Teachers (N = 38)	Experienced Teachers (N = 51)	Difference
Simplified-Enriched	72.1 (10.3)	79.4 (8.5)	7.3
Simplified-Integrated	65.8 (11.7)	71.2 (10.1)	5.4
Simplified-Mixed	64.3 (12.4)	68.7 (11.6)	4.4
Enriched-Integrated	69.7 (10.8)	74.6 (9.2)	4.9
Enriched-Mixed	68.2 (11.1)	72.1 (10.4)	3.9
Integrated-Mixed	71.5 (9.7)	76.8 (8.8)	5.3
<i>Note:</i> Values are presented as Mean (SD). Higher scores indicate greater coherence between the two models.			

Figure 3. Coherence scores across six model pairs for novice (N = 38) and experienced (N = 51) teachers. Error bars represent standard deviations. Values above each pair indicate the difference between experienced and novice teachers' scores. The Simplified-Enriched pair shows the highest coherence for both groups and the largest experience gap (+ 7.3), while the Simplified-Mixed pair shows the lowest coherence for both groups and the smallest gap (+ 4.4). All differences were statistically significant ($p < .05$) except for Enriched-Mixed ($p = .076$). The pattern suggests that professional experience provides the greatest advantage when teachers must strategically adapt content to different ability levels while maintaining conceptual integrity.

Greatest Coherence: Simplified-Enriched Pair — For both groups, the Simplified-Enriched pair exhibited the highest coherence (novice: 72.1; experienced: 79.4). This suggests that teachers find it most natural to maintain coherence when adjusting content along a single dimension of depth and complexity. Experienced teachers showed a 7.3-point advantage, indicating that professional experience is particularly valuable when teachers must strategically select which content to simplify and which to elaborate.

Least Coherence: Simplified-Mixed Pair — The Simplified-Mixed pair showed the lowest coherence (novice: 64.3; experienced: 68.7), representing the greatest challenge: bridging individually differentiated content with collaborative mixed-ability group activity. Experienced teachers showed only a 4.4-point advantage—the smallest difference among all pairs—suggesting that maintaining coherence between individualized and collaborative formats is challenging for all teachers.

Pattern Differences — Experienced teachers demonstrated the greatest advantage over novices in the Simplified-Enriched pair (7.3 points) and the smallest advantage in the Simplified-Mixed pair (4.4 points). The rank order of coherence was consistent across both groups: Simplified-Enriched > Integrated-Mixed > Enriched-Integrated > Enriched-Mixed > Simplified-Integrated > Simplified-Mixed. The **Integrated-Mixed**

pair (novice: 71.5; experienced: 76.8) showed relatively high coherence for both groups, ranking second, as both models share a collaborative, project-based orientation.

4.5. Summary of Key Findings

1. **Overall Coherence Advantage:** Experienced teachers demonstrated significantly higher overall CQ scores compared to novices ($M = 73.8$ vs. 67.4 , $d = 0.62$), indicating that professional experience is meaningfully associated with greater internal coherence across differentiated content.
2. **Dimensional Patterns:** The most pronounced differences were in Conceptual Coherence ($d = 0.79$) and Structural Coherence ($d = 0.77$). Linguistic Coherence showed a non-significant difference ($d = 0.38$), suggesting linguistic adaptation is less dependent on experience, likely due to structured prompt engineering support.
3. **Negative Cases:** Approximately 18% of novice teachers performed above the experienced mean, while about 12% of experienced teachers performed below the novice mean. Exceptional novices demonstrated strong subject knowledge, systematic prompt engineering, and active AI output editing. Low-performing experienced teachers exhibited over-editing, resistance to AI suggestions, and idiosyncratic pedagogy that fragmented coherence.
4. **Overlap Analysis:** The 38.2% overlap in overall CQ distributions suggests substantial common ground between groups. Individual factors beyond experience—such as subject knowledge, prompt engineering skill, and critical engagement with AI—are important determinants of coherence.
5. **Pairwise Patterns:** The Simplified-Enriched pair showed the highest coherence for both groups; the Simplified-Mixed pair showed the lowest. Experienced teachers showed the greatest advantage over novices in the Simplified-Enriched pair (7.3 points) and the smallest in the Simplified-Mixed pair (4.4 points), suggesting that professional experience is most valuable when adapting content to different ability levels while maintaining conceptual integrity.

5. Discussion

The findings of this study offer several contributions to the growing body of research on AI-augmented teaching, teacher professional knowledge, and the nature of coherence in differentiated content generation. This section interprets the key findings, situates them within the broader theoretical landscape, examines their implications, and addresses the critical reflexivity measures that underpin the study's scholarly rigor.

5.1. Interpretation of Key Findings

5.1.1. The Expertise Paradox and the Mediating Role of AI

The finding that experienced teachers demonstrated significantly higher overall CQ scores ($d = 0.62$) aligns with the theoretical expectation that professional experience, through accumulated pedagogical content knowledge (Shulman, 1986, 1987) and refined pedagogical reasoning (Loughran et al., 2016),

enhances teachers' capacity to maintain coherence across differentiated content. The largest group differences in Conceptual Coherence ($d = 0.79$) and Structural Coherence ($d = 0.77$) suggest that experience most strongly shapes teachers' capacity to preserve the conceptual core and maintain logical structure—dimensions of enacted PCK that develop through iterative cycles of planning, teaching, and reflection (Carlson et al., 2019; Bernstein, 1996, 2000).

However, the non-significant difference in Linguistic Coherence ($d = 0.38$, $p = .074$) reveals a crucial nuance: linguistic adaptation is less dependent on professional experience and more accessible to novice teachers, largely because the structured prompt engineering framework (Role + Audience + Format) provided in the workshop appears to have "leveled the playing field" in this dimension. This finding suggests that AI can serve as a "cognitive partner" (Cukurova, 2025) that augments teachers' capacity to produce coherent differentiated content, particularly in dimensions where novice teachers typically struggle. Yet it also raises a cautionary note: Linguistic Coherence may have been higher because it was the dimension most directly influenced by AI output, while Conceptual and Structural Coherence require deeper pedagogical reasoning that AI cannot easily replicate (Dell'Acqua et al., 2026; Loughran et al., 2016).

The negative case analysis further complicates the narrative. Eighteen percent of novice teachers outperformed the experienced mean, while 12% of experienced teachers underperformed the novice mean. Exceptional novice teachers demonstrated strong subject matter knowledge, systematic prompt engineering, and active editing of AI outputs—factors more predictive of high coherence than experience alone. This suggests that AI literacy—the capacity to critically engage with AI-generated content—may be a more important determinant of content quality than years of teaching experience (Chiu et al., 2024; Heung et al., 2026). Conversely, low-performing experienced teachers exhibited "expertise-induced fragmentation"—over-editing AI-generated content and introducing idiosyncratic examples that disrupted coherence (Mishra & Bozkurt, 2026; Ljungqvist et al., 2026). This aligns with the concept of the "jagged technological frontier" (Dell'Acqua et al., 2026), where AI assistance improves performance on some tasks but may not equally support others.

5.1.2. Model Pairs and the Nature of Differentiation

The finding that the Simplified-Enriched pair showed the highest coherence for both groups while the Simplified-Mixed pair showed the lowest coherence provides insight into the cognitive demands of different differentiation strategies. The Simplified-Enriched pair represents "vertical differentiation"—adapting content within the same instructional format to different levels of cognitive demand—which aligns with the core logic of differentiated instruction (Tomlinson, 2014) and may be more intuitive for teachers. The Simplified-Mixed pair, conversely, represents "horizontal differentiation"—bridging fundamentally different instructional formats (individual vs. collaborative). Maintaining coherence across this pedagogical divide requires teachers to coordinate both content and pedagogical decisions.

The differential pattern—experienced teachers showed the greatest advantage in Simplified-Enriched (+ 7.3) and the smallest in Simplified-Mixed (+ 4.4)—suggests that vertical differentiation is where

professional experience provides the most benefit. Experienced teachers' deeper knowledge of content and learning progressions enables more strategic decisions about what to include or exclude at different complexity levels. Horizontal differentiation, however, may require specific knowledge about cooperative learning structures—knowledge less dependent on general teaching experience and more dependent on targeted professional development.

5.1.3. Alternative Hypotheses: Why Novice Teachers Sometimes Outperform Experienced Teachers

The finding that 18.4% of novice teachers performed above the experienced mean supports the alternative hypotheses articulated in Section 2.4. Rather than rehearse the theoretical arguments in full, we focus here on empirical confirmation. First, novice teachers' recent formal training in differentiated instruction and constructivist pedagogies (Day et al., 2006; Pillen et al., 2013) appears to provide explicit frameworks for maintaining conceptual coherence—frameworks that may have been less central in experienced teachers' training (Vanhees et al., 2025). Second, novice teachers adhere more closely to curriculum ideals and instructional templates (Priestley et al., 2015, 2026), avoiding the idiosyncratic modifications that experienced teachers develop over years of practice—modifications that, while reflecting deep contextual knowledge, may introduce coherence-disrupting inconsistencies (Ljungqvist et al., 2026). Third, novice teachers' more fluid professional identity (Pillen et al., 2013; Gao et al., 2025) may make them more adaptable to AI tools, accepting and integrating AI-generated suggestions with less resistance. Finally, the ecological conditions shaping teacher agency (Priestley et al., 2015, 2026) may allow novice teachers, unencumbered by established routines, to exercise greater agency in systematically attending to coherence. These alternative explanations underscore that professional experience is not a unidimensional construct; years of teaching do not automatically translate into better AI-augmented content generation.

5.2. Theoretical Implications

The findings extend existing AI-TPACK frameworks (An et al., 2026; Chen et al., 2026; Sun et al., 2026) by focusing on the *artifactual quality* of teacher-AI collaboration rather than on declarative knowledge about AI. The Consistency Quotient offers a theoretically grounded and empirically validated instrument for assessing this quality. By operationalizing coherence across conceptual, linguistic, and structural dimensions, CQ provides a way to assess whether teachers are effectively using AI to produce content that is not only efficient but also pedagogically coherent—what Carlson et al. (2019) term "enacted PCK."

The findings also have implications for understanding teacher agency in AI-augmented contexts (Priestley et al., 2015, 2026). Experienced teachers' over-editing reflects a form of agency exercised through professional judgment, yet this agency sometimes produced unintended fragmentation. This suggests that teacher agency in AI-augmented contexts requires not only the capacity to make decisions but also the capacity to make *coherence-preserving* decisions—a distinction with significant implications for professional development.

5.3. Practical Implications

Teacher Professional Development

Explicit training in coherence as a quality metric should be integrated into professional development. The high performance of some novice teachers suggests that such training can be effective. Professional development should also address the specific needs of experienced teachers who may over-edit AI outputs, helping them develop "AI pedagogical judgment"—the capacity to evaluate AI outputs and maintain coherence while exercising professional agency.

Preservice Teacher Education

Novice teachers' recent training in differentiation provides a foundation for effective AI collaboration. Preservice programs should build on this foundation by integrating AI literacy into teacher preparation (Chiu et al., 2024; Heung et al., 2026).

AI Tool Design

The finding that Conceptual and Structural Coherence were more challenging suggests that AI tools should explicitly support these dimensions—for example, by flagging potential conceptual inconsistencies or suggesting structural patterns that maintain logical flow. Tools should be designed to support rather than supplant teacher judgment.

5.4. Self-Critique and Reflexivity

In accordance with the study's reflexive commitment, four potential biases are explicitly acknowledged:

Selection of "Coherence"

The choice of coherence as the primary analytical dimension may favor experienced teachers, who have deeper content knowledge and pedagogical reasoning. Future research should examine other dimensions of content quality that might reveal different patterns.

IAT Framework

The study was framed within the Integrated AI Triad (IAT) model, which may have influenced the analysis by prioritizing coherence as a key dimension of teacher competence. Future research should test generalizability across different theoretical lenses.

AI-Assisted Analysis

The coding process included AI-assisted thematic mapping, which may have introduced systemic biases (Messner et al., 2025). To mitigate this, a double-coding protocol was implemented with human coders, and disagreements were resolved through consensus discussion. The coding framework was explicitly grounded in established theoretical traditions (Bernstein, 1996, 2000; Mayer, 2024; van Es & Sherin, 2026).

Mitigation Steps: Four critical reflexivity measures were incorporated: (1) alternative hypotheses examination (Section 2.4), (2) negative case analysis (Section 4.2), (3) overlap analysis (Section 4.3), and (4) self-critique in the discussion, explicitly acknowledging how the selection of coherence may favor experienced teachers and how AI-assisted analysis may have introduced biases.

6. Conclusion

This study examined the internal coherence of AI-generated educational content produced by novice and experienced teachers within the Integrated AI Triad (IAT) framework. Through comparative content analysis of 356 content samples generated by 89 teachers across four differentiated models, the study introduced the Consistency Quotient (CQ) as a composite measure of conceptual, linguistic, and structural coherence.

The findings reveal that experienced teachers demonstrate significantly higher overall CQ scores ($M = 73.8$ vs. 67.4 , $d = 0.62$), with the most pronounced differences in Conceptual Coherence ($d = 0.79$) and Structural Coherence ($d = 0.77$). Linguistic Coherence showed a non-significant difference ($d = 0.38$), suggesting that structured prompt engineering frameworks can effectively support linguistic adaptation regardless of experience. However, the 38.2% overlap between distributions and the identification of negative cases—18.4% of novice teachers outperforming the experienced mean and 11.8% of experienced teachers underperforming the novice mean—indicate that professional experience is not a deterministic predictor of coherence quality. Individual factors such as subject matter knowledge, prompt engineering skill, and critical engagement with AI outputs are important determinants.

Pairwise model analysis revealed that the Simplified-Enriched pair showed the highest coherence for both groups, while the Simplified-Mixed pair showed the lowest, suggesting that vertical differentiation (adjusting content complexity) is more cognitively accessible than horizontal differentiation (bridging individualized and collaborative formats). Experienced teachers showed the greatest advantage in the Simplified-Enriched pair (+ 7.3) and the smallest in the Simplified-Mixed pair (+ 4.4), indicating that professional experience is most valuable when teachers must strategically adapt content while maintaining conceptual integrity.

6.1. Contributions

This study makes three significant contributions to the literature on AI-augmented teaching. First, it introduces the Consistency Quotient (CQ) as a novel, empirically grounded instrument for assessing the quality of teacher-AI collaboration, moving beyond efficiency metrics to capture pedagogical integrity. The CQ offers a theoretically grounded framework operationalizing enacted PCK (Carlson et al., 2019) in AI-augmented contexts. Second, it provides empirical evidence on how professional experience shapes teachers' capacity to maintain coherence when generating differentiated content with GenAI, offering insights into the developmental trajectories of AI-TPACK. Third, it extends existing AI-TPACK frameworks by focusing on the artifactual quality of teacher-edited AI outputs, thereby bridging the gap between

teacher knowledge and instructional practice. The study's four-part reflexive structure—Alternative Hypotheses Examination, Negative Case Analysis, Overlap Analysis, and Self-Critique—ensures scholarly rigor and mitigates confirmation bias.

6.2. Limitations

Several limitations warrant consideration. First, the study was conducted with Iranian teachers, and cultural, curricular, and contextual factors may limit generalizability to other educational systems. Second, the data were generated in a professional development workshop context, not authentic classroom practice, where time constraints, resource availability, and instructional demands may differ significantly. Third, the study focused on a single content domain per teacher, and patterns of coherence may vary across subject areas. Fourth, the study did not measure actual student learning outcomes, leaving open the question of whether higher CQ scores translate to improved student understanding. Fifth, the data were self-reported (teacher-generated assignments), not observed classroom instruction. Sixth, the researcher's dual role as IAT developer and workshop instructor may have influenced the analysis despite the reflexive measures incorporated.

6.3. Future Research

Future research should examine the generalizability of these findings across different cultural contexts and educational systems. Cross-cultural replication studies would assess whether the observed patterns hold in diverse settings. Studies linking CQ scores to student learning outcomes would establish the practical significance of coherence quality. Longitudinal designs tracking teachers' CQ development over time would illuminate the trajectory of AI-TPACK growth. Qualitative studies investigating teachers' decision-making processes during AI-augmented content generation would provide deeper insight into how coherence is maintained or disrupted. Additionally, research comparing teachers with different levels of AI literacy training would help disentangle the relative contributions of professional experience and AI-specific training. Finally, intervention studies testing professional development programs designed to enhance coherence quality would provide evidence for the malleability of CQ.

References

1. Afidah NT, Hernawan AH (2025) Understanding by Design (UbD) Learning Model for Developing Independent Learning Competencies in a Madrasah Ibtidaiyah. *Khazanah Pendidikan Islam* 7(2):183–194
2. An X, Bai J, Li Y, Zhou Y, Chai CS, Hong J-C (2026) Development of an Artificial Intelligence Technology Pedagogical Content Knowledge (AI-TPACK) Scale for Human and AI Concurrently as Teachers. *Asia-Pacific Educ Researcher* 35(3):483–493
3. Bernstein B (1996) *Pedagogy, symbolic control, and identity: Theory, research, critique*. Taylor & Francis
4. Bernstein B (2000) *Pedagogy, symbolic control, and identity*, vol 5. Rowman & Littlefield

5. Biesta G, Priestley M, Robinson S (2015) The role of beliefs in teacher agency. *Teachers Teach* 21(6):624–640
6. Bosch N, Härkki T, Seitamaa-Hakkarainen P (2025) Teachers as reflective learning experience designers: Bringing design thinking into school-based design and maker education. *Int J Child-Comput Interact* 43:100695
7. Braun V, Clarke V (2020) One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Res Psychol* 18(3):328–352
8. Braun V, Clarke V (2023) Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychol Rev* 17(4):695–718
9. Cao J, Wang S, Chen X, Schunn C (2026) Co-develop Learning Tasks with Generative AI: Mapping Challenges, Gains, and Professional Development Needs. *IEEE Trans Learn Technol*
10. Carlson J, Daehler KR, Alonzo AC, Barendsen E, Berry A, Borowski A, Wilson CD (2019) The refined consensus model of pedagogical content knowledge in science education. Repositioning pedagogical content knowledge in teachers' knowledge for teaching science. Springer, pp 77–94
11. Chan CKY, Lee KK (2025) The balancing act between AI and authenticity in assessment: A case study of secondary school students' use of GenAI in reflective writing. *Comput Educ*, 105399
12. Chen J, Loke SP, Teh SY (2026) From TPACK to I-TPACK: exploring teachers' perceptions and experiences with AI tools in pedagogical practice in Chinese higher education. *Interact Learn Environ*, 1–20
13. Chiu TKF, Ahmad Z, Ismailov M, Sanusi IT (2024) What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers Education: Artif Intell* 6:100171
14. Chiu TKF, Xia Q, Zhou X, Chai CS, Cheng M (2023) Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers Education: Artif Intell* 4:100118
15. Corbin T, Dawson P, Nicola-Richmond K, Partridge H (2025) Where's the line? It's an absurd line': towards a framework for acceptable uses of AI in assessment. *Assess Evaluation High Educ* 50(5):705–717
16. Cukurova M (2024) The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *Br J Edu Technol* 55(6):2412–2428
17. Cukurova M (2025) Promoting and Protecting Teacher Agency in the Age of Artificial Intelligence. UNESCO Publishing
18. Day C, Kington A, Stobart G, Sammons P (2006) The personal and professional selves of teachers: stable and unstable identities. *Br Edu Res J* 32(4):601–616
19. Dell'Acqua F, McFowland III, Mollick E, Lifshitz E, Kellogg H, Rajendran KC, Lakhani S, K. R (2026) Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organ Sci* 37(2):403–423

20. Dinçer B (2026) Balancing artificial intelligence and human guidance in teacher education: A quasi-experimental study. *Sage Open* 16(1):21582440261430067
21. Dogan S (2025) A human-AI hybrid co-design model: Planning effective instruction with ChatGPT. *Asian J Distance Educ*, 20(1)
22. Fernández-Sánchez A, Lorenzo-Castiñeiras JJ, Sánchez-Bello A (2025) Navigating the future of pedagogy: The integration of AI tools in developing educational assessment rubrics. *Eur J Educ*, 60(1), e12826
23. Festinger L (1957) A theory of cognitive dissonance. Stanford University Press. [引用形式 著者 年]
24. Gao Y, Wang X, Reynolds BL (2025) The interplay of emotion regulation, resilience, and teacher identity: A structural equation modeling study of Chinese middle school English teachers. *Psychol Sch* 62(9):2876–2887
25. Heung YME, Yin HB, English L, Chiu TKF (2026) Theorizing AI Literacy Development using Habermas' Three Cognitive Knowledge Interests from a Systematic Review: A STEM Interdisciplinary Perspective. *Comput Educ*, 105492
26. Hobbs L (2020) Learning to teach science out-of-field: A spatial-temporal experience. *J Sci Teach Educ* 31(7):725–745
27. Howard GS (1980) Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Eval Rev* 4(1):93–106 [引用形式 著者 年]
28. Kaswa A (2025) The Evolution of Cognitive Dissonance: A Review of Foundational Paradigms, Theoretical Refinements, and Contemporary Directions. *Basic Appl Soc Psychol*, 1–5
29. Koehler M, Mishra P (2009) What is technological pedagogical content knowledge (TPACK)? *Contemp Issues Technol Teacher Educ* 9(1):60–70
30. Laurillard D (2024) Trust the teachers: A collaborative approach to learning design solutions. *Envisioning the future of education through design*. Springer, pp 119–145
31. Ljungqvist M, Sonesson A, Selwyn N (2026) Fixing teachers' problems? exploring teachers' repair and maintenance work around generative AI technologies. *Discourse: Stud Cult Politics Educ*, 1–12
32. Loughran J, Keast S, Cooper R (2016) Pedagogical reasoning in teacher education. *International Handbook of Teacher Education*. Springer, pp 387–421
33. Luckin R, Holmes W (2016) Intelligence unleashed: An argument for AI in education. Pearson
34. Mayer RE (2024) The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychol Rev* 36(1):8
35. Messner R, Smith S, Richards C (2025) Artificial intelligence and qualitative data analysis: Epistemological incongruences and the future of the human experience. *Int J Qualitative Methods* 24:16094069251371481
36. Mishra P, Bozkurt A (2026) TPACK in the Age of Context and Artificial Intelligence: Punya Mishra on the Wicked Problem of AI and Why Algorithms Can't Replace Context. *Open Praxis*, 18(2)

37. Pillen M, Beijaard D, Brok PD (2013) Tensions in beginning teachers' professional identity development, accompanying feelings and coping strategies. *Eur J Teacher Educ* 36(3):240–260
38. Priestley M, Biesta GJ, Philippou S, Robinson S (2015) The teacher and the curriculum: Exploring teacher agency. *The SAGE Handbook of Curriculum, Pedagogy and Assessment*, 27
39. Priestley M, Biesta G, Robinson S (2026) *Teacher agency: An ecological approach*, 2nd edn. Bloomsbury Publishing
40. Ritzhaupt AD, Dawson K, Crompton H (2026) Toward guidance and principled values on the ethical, responsible, and transparent use of GenAI in the educational research process. *J Res Technol Educ* 58(2):235–249
41. Rosyidi AH, Sa'dijah C, Purnomo H, Abdullah AH (2025) From complex to simple: Analyzing prospective teachers' analogy reasoning in creating accessible geometry problems. *Kalamatika: Jurnal Pendidikan Matematika* 10(2):73–93
42. Samuel A, Soh M, Jung E (2025) Enhancing reflective practice with ChatGPT: a new approach to assignment design. *Med Teach* 47(9):1438–1440
43. Selwyn N (2019) *Should robots replace teachers? AI and the future of education*. Wiley
44. Selwyn N (2021) *Education and technology: Key issues and debates*, 3rd edn. Bloomsbury Publishing
45. Shulman LS (1986) Those who understand: Knowledge growth in teaching. *Educational Researcher* 15(2):4–14
46. Shulman L (1987) Knowledge and teaching: Foundations of the new reform. *Harv Educational Rev* 57(1):1–23
47. Sun Y, Xin H, Li S, Niu Q, Chai CS, Huang L, Chen G (2026) Modeling AI-TPACK in Practice Insights from Teachers Multi-Agent Workflow Design. *arXiv preprint arXiv:2605.13906*
48. Sweller J (1988) Cognitive load during problem solving: Effects on learning. *Cogn Sci* 12(2):257–285
49. Talebzadeh H (2025) *The Integrated AI Triade (IAT): A Systemic Framework for Redefining the Roles of Teachers, Managers, and Families in the AI Era*. Zenodo. <https://doi.org/10.5281/zenodo.17914437>
50. Talebzadeh H (2025) *Bilingual Teacher Guide: Session 2 - Personalized Content Creation*. Zenodo. <https://doi.org/10.5281/zenodo.17536752>
51. Tomlinson CA (2014) *The differentiated classroom: Responding to the needs of all learners*, 2nd edn. ASCD
52. van Es EA, Sherin MG (2026) Exploring teacher noticing in practice: the role of shaping. *ZDM – Math Educ* 58:19–31
53. Vanhees C, Nijlunsing J, Muijs D, Crato N, Wils M, Wiliam D, Kirschner PA (2025) The role of knowledge-rich curricula in promoting deep thinking and complex skill acquisition. *Learn Individual Differences* 121:102729
54. Wang L (2026) Within-teacher contrasts of in-field and out-of-field noticing: Integrated knowledge and teaching experience. *Teach Teacher Educ* 172:105671

55. Zawacki-Richter O, Cefa B, Bai JYH (2025) Towards reproducible systematic reviews in Open, Distance, and Digital Education—An umbrella mapping review. *Rev Educ*, 13(1), e70031

56. Zhou X, Li Y, Chai CS, Chiu TKF (2025) Defining, enhancing, and assessing artificial intelligence literacy and competency in K-12 education from a systematic review. *Interact Learn Environ* 33(10):5766–5788

Figures

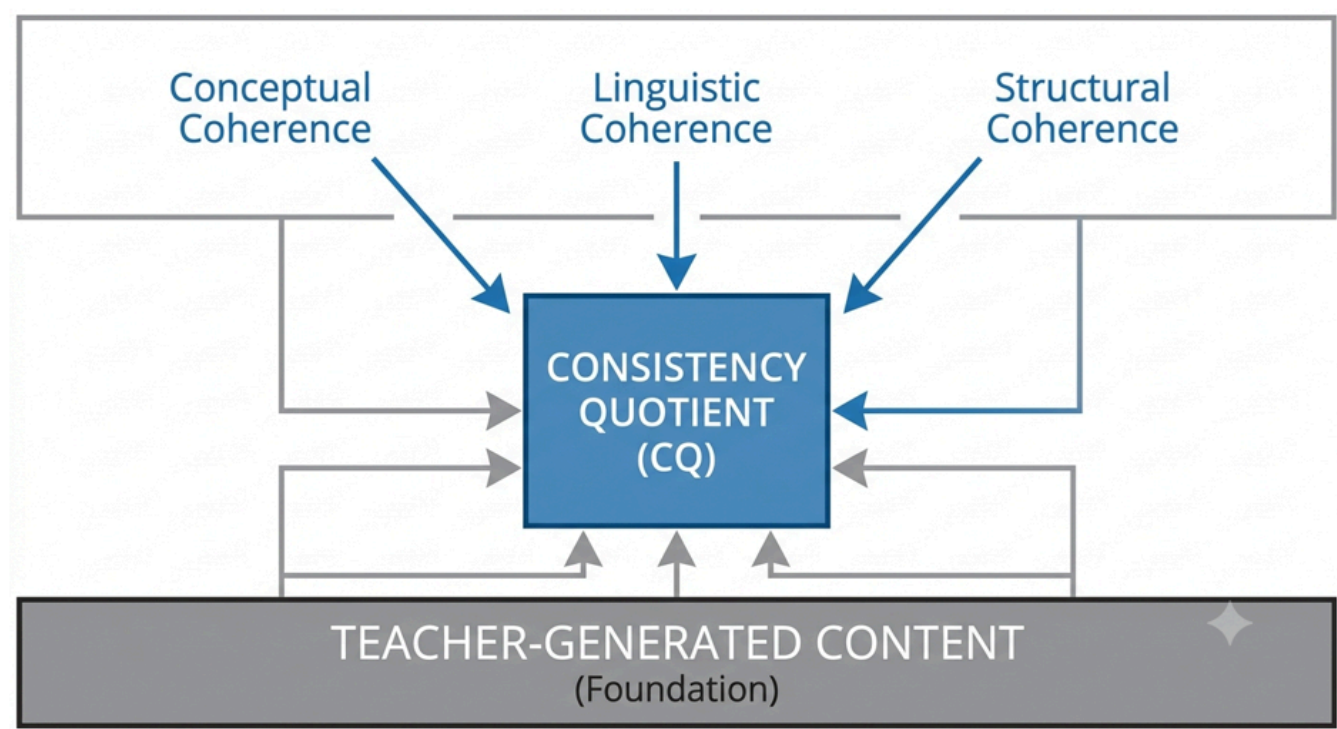


Figure 1

Conceptual Model of Coherence Dimensions and Consistency Quotient

Figure 2. Distribution of Consistency Quotients by Teacher Experience Level

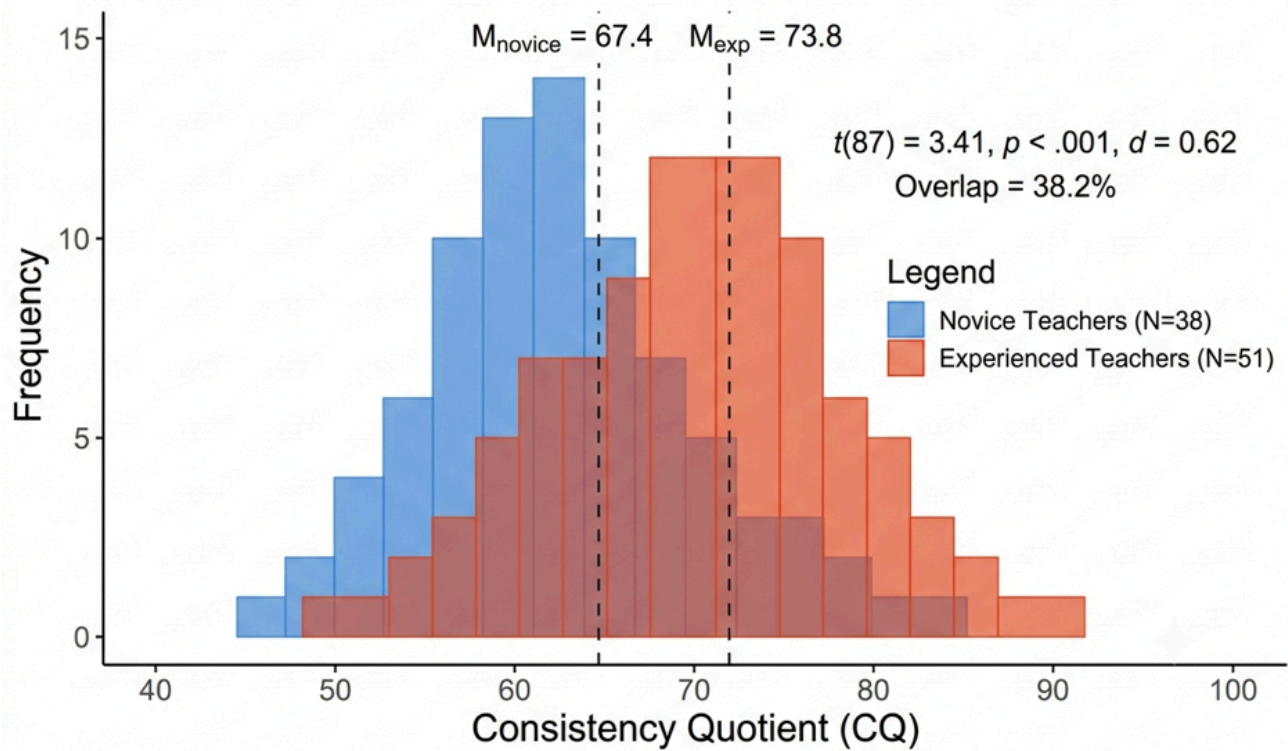


Figure 2

Distribution of Consistency Quotients by Teacher Experience Level 4.1.2. Dimensional Coherence Analysis

Distribution of Consistency Quotient (CQ) scores for novice (N=38) and experienced (N=51) teachers. Vertical dashed lines indicate group means ($M_{\text{novice}} = 67.4$, $M_{\text{exp}} = 73.8$). The overlap between distributions is 38.2%. The difference is statistically significant ($t(87) = 3.41$, $p < .001$, Cohen's $d = 0.62$).

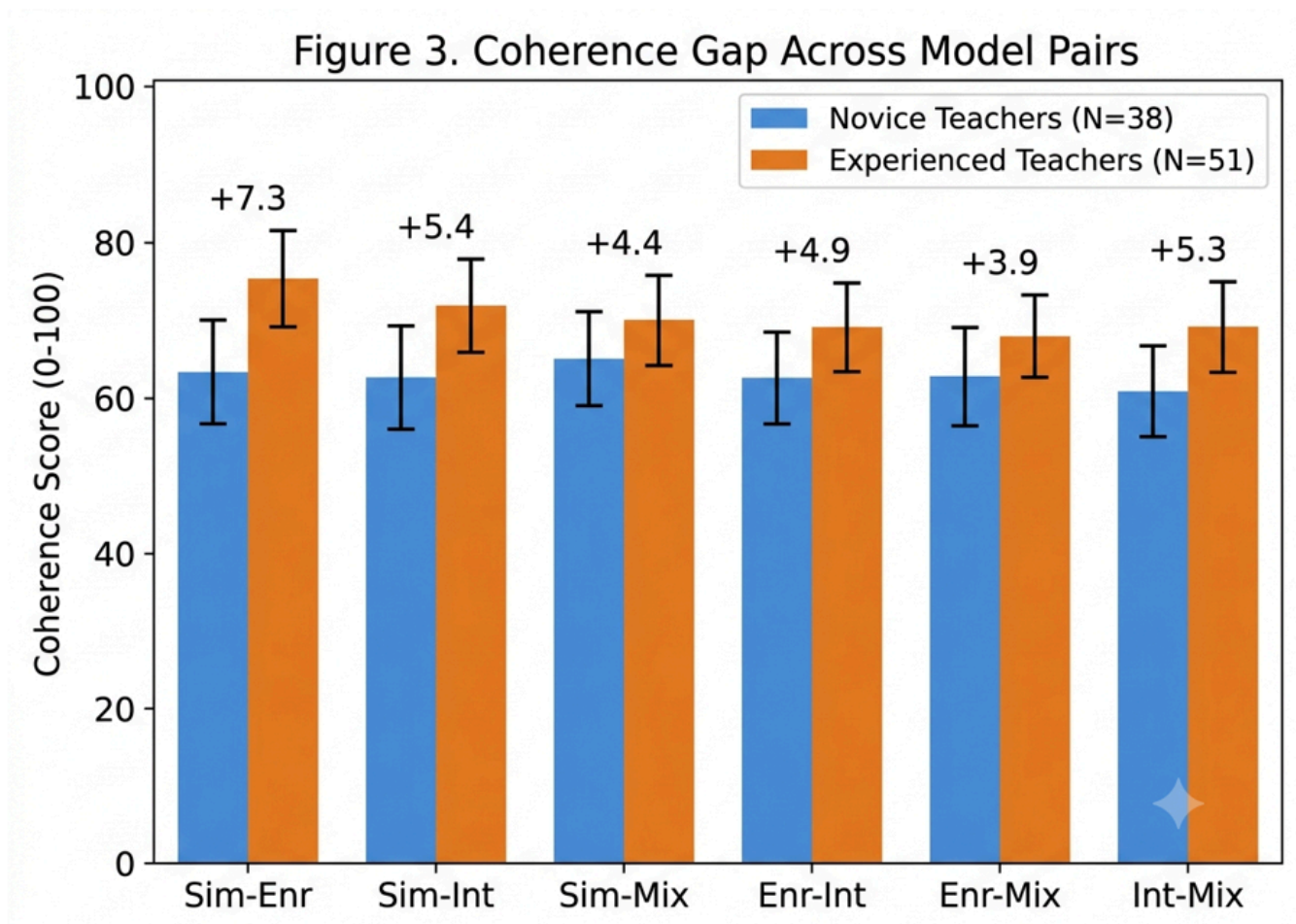


Figure 3

Coherence Gap Across Model Pairs – Bar Chart

Coherence scores across six model pairs for novice (N=38) and experienced (N=51) teachers. Error bars represent standard deviations. Values above each pair indicate the difference between experienced and novice teachers' scores. The Simplified-Enriched pair shows the highest coherence for both groups and the largest experience gap (+7.3), while the Simplified-Mixed pair shows the lowest coherence for both groups and the smallest gap (+4.4). All differences were statistically significant ($p < .05$) except for Enriched-Mixed ($p = .076$). The pattern suggests that professional experience provides the greatest advantage when teachers must strategically adapt content to different ability levels while maintaining conceptual integrity.